Robust Statistics on Riemannian Manifolds via the Geometric Median

P. Thomas Fletcher
School of Computing
University of Utah
fletcher@sci.utah.edu

Suresh Venkatasubramanian
School of Computing
University of Utah
suresh@cs.utah.edu

Sarang Joshi
Biomedical Engineering
University of Utah
sjoshi@sci.utah.edu

Abstract

The geometric median is a classic robust estimator of centrality for data in Euclidean spaces. In this paper we formulate the geometric median of data on a Riemannian manifold as the minimizer of the sum of geodesic distances to the data points. We prove existence and uniqueness of the geometric median on manifolds with non-positive sectional curvature and give sufficient conditions for uniqueness on positively curved manifolds. Generalizing the Weiszfeld procedure for finding the geometric median of Euclidean data, we present an algorithm for computing the geometric median on an arbitrary manifold. We show that this algorithm converges to the unique solution when it exists. This method produces a robust central point for data lying on a manifold, and should have use in a variety of vision applications involving manifolds. We give examples of the geometric median computation and demonstrate its robustness for three types of manifold data: the 3D rotation group, tensor manifolds, and shape spaces.

1. Introduction

Riemannian geometry plays a central role in many computer vision applications, in particular, the study of shape variability. Recently there has been substantial interest in the statistical characterization of data that are best modeled as elements of a Riemannian manifold, rather than as points in Euclidean space [14, 20, 28, 32]. In previous work [7, 14, 28], the notion of centrality of empirical data was defined via the Fréchet mean [15], which was first developed for manifold-valued data by Karcher [18]. Although the mean is an obvious central representative, one of its major drawbacks is its lack of robustness, i.e., it is sensitive to outliers.

Robust statistical estimation in Euclidean spaces is now a field in its own right, and numerous robust estimators exist. However, no such robust estimators have been proposed for data lying on a manifold. One of the most common robust estimators of centrality in Euclidean spaces is the *geomet-ric median*. Although the properties of this point have been extensively studied since the time of Fermat, (this point is often called the Fermat-Weber point), no generalization of this estimator exists for manifold-valued data. In this paper we extend the notion of geometric median to general Riemannian manifolds, thus providing a robust statistical estimator of centrality for manifold-valued data. We prove some basic properties of the generalization and exemplify its robustness for data on common manifolds encountered in computer vision. In this paper we are particularly interested in the statistical characterization of shapes given an ensemble of empirical measurements. Although the methods presented herein are quite general, for concreteness we will focus on the following explicit examples: i) the space of 3D rotations, ii) the space of positive-definite tensors, and iii) the space of planar shapes.

2. Outliers, Robust Estimators And The Geometric Median

Robust estimation has received considerable attention in the computer vision literature (see [25, 33] for a review). Several robust estimators proposed in the statistics literature, such as M-estimators and least median of squares [17], have been used in computer vision. Other robust techniques, such as the Hough transform and random sample consensus (RANSAC) [12], have been developed within computer vision. Applications of robust estimators include image segmentation, image feature detection, surface geometry from range data, point correspondences and fundamental matrix estimation, among others.

Outliers in data can throw off estimates of centrality based on the mean. One possible solution to this problem is outlier deletion, but removing outliers often merely promotes other data points to outlier status, forcing a large number of deletions before a reliable low-variance estimate can be found. The theory of robust estimators formalizes the idea that no individual point should affect measures of central tendency. The measure of robustness of an estimator is the *breakdown point*; formally, it is the fraction of

the data that can be "dragged to infinity" (i.e., completely corrupted) without affecting the boundedness of the estimator. Clearly, the mean, whether it be a standard centroid or the more general Fréchet mean, has a breakdown point of 0, since as any single data point is dragged to infinity, the mean will grow without bound.

The theory of robust estimation has led to the development of numerous robust estimators, of which the L_1-estimator, also known as the geometric median, is one of the best known. Given a set of points $\{x_i, i = 1, \cdots, n\} \in \mathbb{R}^d$, with the usual Euclidean norm $\|x\|$, the L_1-estimator is defined as the point $m \in \mathbb{R}^d$ minimizing $\sum_{i=1}^n \|m - x_i\|$. It can be shown [24] that this estimator has a breakdown point of 0.5, which means that *half of the data* needs to be corrupted in order to corrupt this estimator. In Figure 1 we illustrate this by showing how the geometric median and the mean are displaced in the presence of a few outliers.
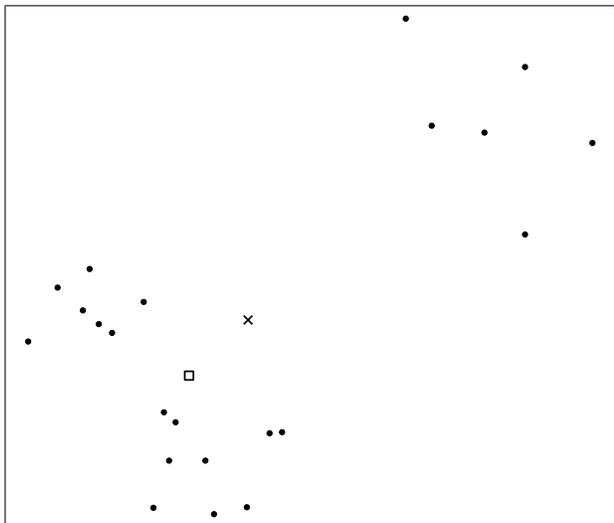

Figure 1. The geometric median (marked with a □) and mean (marked with a ×) for a collection of points in the plane. Notice how the few outliers at the top right of the picture have forced the mean away from the points, whereas the median remains centrally located.

The existence and uniqueness of the the median in $\mathbb{R}^d$ follows directly from the convexity of the distance function. In one dimension, the geometric median is the point that divides the point set into equal halves on either side (if n is odd) and is any point on the line segment connecting the two middle points (if n is even). In general however, computing the geometric median is difficult; Bajaj has shown that the solution cannot be expressed using radicals (arithmetic operations, and k^{th} roots) [1].

There are two main approaches to computing the geometric median of a collection of points in $\mathbb{R}^d$. One way is to compute an *approximate* median $\tilde{m}$ such that $\sum_{i=1}^n \|\tilde{m} - x_i\|$ is at most a $(1+\epsilon)$-factor larger than cost of

the optimal median. This can be computed using the ellipsoid method [8]. A more efficient algorithm achieving the same result is due to Bose *et al.* [6].

These algorithms do not generalize beyond Euclidean spaces. A more general iterative algorithm due to Weiszfeld [34] and later improved by Kuhn and Kuenne [22] and Ostresh [27] converges to the optimal solution in Euclidean spaces [21], and was subsequently been generalized to Banach spaces by Eckhardt [11].

3. The Riemannian Geometric Median

Let M be a Riemannian manifold. Given points $x_1, \ldots, x_N \in M$ and corresponding positive real weights $w_i, \ldots, w_N$, with $\sum_i w_i = 1$, define the weighted sum-of-distances function $f(x) = \sum_i w_i d(x, x_i)$, where d is the Riemannian distance function on M. Throughout, we will assume that the x_i lie in a convex set $U \subset M$, i.e., any two points in U are connected by a unique shortest geodesic lying entirely in U. We define the *weighted geometric median*, m, as the minimizer of f, i.e.,

$$m = \arg \min_{x \in M} \sum_i^N w_i d(x, x_i). \tag{1}$$

When all the weights are equal, $w_i = 1/N$, we call m simply the *geometric median*. In contrast, the Fréchet mean, or Karcher mean [18], of a set of points on a Riemannian manifold is defined, via the generalization of the least squares principle in Euclidean spaces, as the minimizer of the sum-of-squared distances function,

$$\mu = \arg \min_{x \in M} \sum_i^N w_i d^2(x, x_i). \tag{2}$$

We begin our exploration of the geometric median with a discussion of the Riemannian distance function. Given a point $p \in M$ and a tangent vector $v \in T_p M$, where $T_p M$ is the tangent space of M at p, there is a unique geodesic, $\gamma : [0, 1] \to M$, starting at p with initial velocity v. The Riemannian exponential map, $\text{Exp}_p : T_p M \to M$, maps the vector v to the endpoint of this geodesic, i.e., $\text{Exp}_p(v) = \gamma(1)$. The exponential map is locally diffeomorphic onto a neighborhood of p. Let $V(p)$ be the largest such neighborhood. Then within $V(p)$ the exponential map has an inverse, the Riemannian log map, $\text{Log}_p : V(p) \to T_p M$. For any point $q \in V(p)$ the Riemannian distance function is given by $d(p, q) = \|\text{Log}_p(q)\|$. For a fixed point $p \in M$, the gradient of the Riemannian distance function is $\nabla_x d(p, x) = -\text{Log}_x(p)/\|\text{Log}_x(p)\|$ for $x \in V(p)$. Notice that this is a unit vector at x, pointing away from p (compare to the Euclidean distance function).

The diameter of U, denoted $\text{diam}(U)$, is the maximal distance between any two points in U. Using the convexity

properties of the Riemannian distance function (see the Appendix for more details), we have the following existence and uniqueness result for the geometric median.

Theorem 1. *The weighted geometric median defined by* (1) *exists and is unique if (a) the sectional curvatures of M are nonpositive, or if (b) the sectional curvatures of M are bounded above by $\Delta > 0$ and $\mathrm{diam}(U) < \pi/(2\sqrt{\Delta})$.*

Proof. Let $\gamma : [a,b] \to U$ be a geodesic. By the arguments in the Appendix, the distance function to any x_i is convex, that is, $(d^2/dt^2)d(x_i, \gamma(t)) \geq 0$. Since the weighted sum-of-distances function $f(x)$ is a convex combination of such functions, it is also convex. Furthermore, since the x_i do not all lie on the same geodesic, the vector $\mathrm{Log}_{\gamma(t)}(x_k)$ is not tangential to $\gamma(t)$ for at least one $k \in [1, N]$. Therefore, by Lemma 1 we have $(d^2/dt^2)d(x_k, \gamma(t)) > 0$, and $f(x)$ is a *strictly* convex function, which implies that the minimization (1) has a unique solution. □

An isometry of a manifold M is a diffeomorphism ϕ that preserves the Riemannian distance function, that is, $d(x,y) = d(\phi(x), \phi(y))$ for all $x, y \in M$. The set of all isometries forms a Lie group, called the isometry group. It is clear from the definition of the geometric median (1) that the geometric median is invariant under the isometry group of M. In other words, if m is the geometric median of $\{x_i\}$ and ϕ is an isometry, then $\phi(m)$ is the geometric median of $\{\phi(x_i)\}$. This is a property that the geometric median shares with the Fréchet mean.

4. The Weiszfeld Algorithm for Manifolds

For Euclidean data the geometric median can be computed by an algorithm introduced by Weiszfeld [34] and later improved by Kuhn and Kuenne [22] and Ostresh [27]. The procedure iteratively updates the estimate m_k of the geometric median using essentially a steepest descent on the weighted sum-of-distances function, f. For a point $x \in \mathbb{R}^n$ not equal to any x_i, the gradient of f exists and is given by

$$\nabla f(x) = \sum_{i=1}^{N} w_i(x - x_i)/\|x - x_i\|. \qquad (3)$$

The gradient of $f(x)$ is not defined at the data points $x = x_i$. The iteration for computing the geometric median due to Ostresh is

$$m_{k+1} = m_k - \alpha G_k,$$
$$G_k = \sum_{i \in I_k} \frac{w_i x_i}{\|x_i - m_k\|} \cdot \left(\sum_{i \in I_k} \frac{w_i}{\|x_i - m_k\|} \right)^{-1}, \quad (4)$$

where $I_k = \{i \in [1, N] : m_k \neq x_i\}$, and $\alpha > 0$ is a step size. Notice if the current estimate m_k is located at a data

point x_i, then this term is left out of the summation because the distance function is singular at that point. Ostresh [27] proves that the iteration in (4) converges to the unique geometric median for $0 \leq \alpha \leq 2$ and when the points are not all colinear. This follows from the fact that f is strictly convex and (4) is a contraction, that is, $f(m_{k+1}) < f(m_k)$ if m_k is not a fixed point.

Now for a general Riemannian manifold M, the gradient of the Riemannian sum-of-distances function is given by

$$\nabla f(x) = -\sum_{i=1}^{N} w_i \, \mathrm{Log}_x(x_i)/d(x, x_i), \qquad (5)$$

where again we require that $x \in U$ is not one of the data points x_i. This leads to a natural steepest descent iteration to find the Riemannian geometric median, analogous to (4),

$$m_{k+1} = \mathrm{Exp}_{m_k}(\alpha v_k),$$
$$v_k = \sum_{i \in I_k} \frac{w_i \, \mathrm{Log}_{m_k}(x_i)}{d(m_k, x_i)} \cdot \left(\sum_{i \in I_k} \frac{w_i}{d(m_k, x_i)} \right)^{-1}. \quad (6)$$

The following result for positively curved manifolds shows that this procedure converges to the unique weighted geometric median when it exists.

Theorem 2. *If the sectional curvatures of M are nonnegative and the conditions (b) of Theorem 1 are satisfied, then $\lim_{k \to \infty} m_k = m$ for $0 \leq \alpha \leq 2$.*

Proof. We use the fact that the Euclidean Weiszfeld iteration, given by (4), is a contraction. First, define $\tilde{f}(v) = \sum_i w_i \|v - \mathrm{Log}_{m_k}(x_i)\|$, i.e., $\tilde{f}$ is the weighted sum-of-distances function for the log-mapped data, using distances in $T_{m_k}M$ induced by the Riemannian norm. Notice that the tangent vector v_k defined in (6) is exactly the same computation as the Euclidean Weiszfeld iteration (4), replacing each x_i with the tangent vector $\mathrm{Log}_{m_k}(x_i)$. Therefore, we have the contraction property $\tilde{f}(\alpha v_k) < \tilde{f}(0)$. However, geodesics on positively curved manifolds converge, which means that distances between two points on the manifold are closer than their images under the log map. (This is a direct consequence of the Toponogov Comparison Theorem, see [9]). In other words, $d(\mathrm{Exp}_{m_k}(\alpha v_k), x_i) < \|\alpha v_k - \mathrm{Log}_{m_k}(x_i)\|$. This implies that $f(m_{k+1}) = f(\mathrm{Exp}_{m_k}(\alpha v_k)) < \tilde{f}(\alpha v_k) < \tilde{f}(0) = f(m_k)$. (The last equality follows from $\|\mathrm{Log}_{m_k}(x_i)\| = d(m_k, x_i)$.) Therefore, (6) is a contraction, which combined with f being strictly convex, proves that it converges to the unique solution m. □

We believe that a similar convergence result will hold for negatively curved manifolds as well (with an appropriately chosen step size α). Our experiments presented in the next

section for tensor data (Section 5.2) support this belief. The tensor manifold has nonpositive curvature, and we found the procedure in (6) converged for $\alpha = 1$. Proving convergence in this case is an area of future work.

5. Applications

In this section we present results of the Riemannian geometric median computation on 3D rotations, symmetric positive-definite tensors, and planar shapes. For each example the geometric median is computed using the iteration presented in Section 4, which only requires computation of the Riemannian exponential and log maps. Therefore, the procedure is applicable to a wide class of manifolds used in computer vision beyond those presented here. The Fréchet mean is also computed for comparison using a gradient descent algorithm as described in [28] and elsewhere. It is important to note that unlike the Euclidean case where the mean can be computed in closed-form, both the Fréchet mean and geometric median computations for general manifolds are iterative, and we did not find any appreciable difference in the computation times in the examples described below.

5.1. Rotations

We represent 3D rotations as the unit quaternions, $\mathbb{H}_1$. A quaternion is denoted as $q = (a, v)$, where a is the "real" component and $v = bi + cj + dk$. Geodesics in the rotation group are given simply by constant speed rotations about a fixed axis. Let $e = (1, 0)$ be the identity quaternion. The tangent space $T_e\mathbb{H}_1$ is the vector space of quaternions of the form $(0, v)$. The tangent space at an arbitrary point $q \in \mathbb{H}_1$ is given by right multiplication of $T_e\mathbb{H}_1$ by q. The Riemannian exponential map is $\mathrm{Exp}_q((0, v) \cdot q) = (\cos(\theta/2), v \cdot \sin(\theta/2)/\theta) \cdot q$, where $\theta = \|v\|$. The log map is given by $\mathrm{Log}_q((a, v) \cdot q) = (0, \theta v/\|v\|) \cdot q$, where $\theta = 2\arccos(a)$.

To demonstrate the geometric median computations for 3D rotations, we generated a random collection of 20 quaternions. First, random tangent vectors were sampled from an isotropic Gaussian distribution with $\mu = 0, \sigma = \pi/30$ in the tangent space at the identity. Next, the exponential map was applied to these random tangent vectors to produce random elements of $\mathbb{H}_1$, centered about the identity. The same procedure was repeated to generate sets of 5, 10, and 15 random outliers, whose mean now was rotated by 90 degrees from the original set. A sample of 10 of the original random rotations are displayed as 3D frames in the top-left row of Figure 2 along with 10 of the outliers in the bottom-left row.

We computed both the Fréchet mean and the geometric median of the original rotation dataset with 0, 5, 10, and 15 outliers included. This corresponds to an outlier percentage

of 0%, 20%, 33%, and 43%, respectively. The geometric median was computed using the iteration in (6). The Fréchet mean was computed using the gradient descent algorithm described in [7]. Both algorithms converged in under 10 iterations in a fraction of a second for all cases. The results are shown in the right column of Figure 2. The geometric median remains relatively stable even up to an addition of 15 outliers. In contrast, the Fréchet mean is dragged noticeably towards the outlier set.

5.2. Tensors

Positive definite symmetric matrices, or tensors, have a wide variety of uses in computer vision and image analysis, including texture analysis, optical flow, image segmentation, and neuroimage analysis. The space of positive definite symmetric tensors has a natural structure as a Riemannian manifold. Manifold techniques have successfully been used in a variety of applications involving tensors, which we briefly review now.

Diffusion tensor magnetic resonance imaging (DT-MRI) [2] gives clinicians the power to image *in vivo* the structure of white matter fibers in the brain. A 3D diffusion tensor models the covariance of the Brownian motion of water at a voxel, and as such is required to be a 3×3, symmetric, positive-definite matrix. Recent work [3, 13, 29] has focused on Riemannian methods for statistical analysis (Fréchet means and variability) and image processing of diffusion tensor data. The structure tensor [4] is a measure of edge strength and orientation in images and has found use in texture analysis and optical flow. Recently, Rathi *et al.* [30] have used the Riemannian structure of the tensor space for segmenting images. Finally, the Riemannian structure of tensor space has also found use in the analysis of structural differences in the brain, via tensor based morphometry [23].

We briefly review the differential geometry of tensor manifolds, which is covered in more detail in [3, 13, 29]. Recall that a real $n \times n$ matrix A is symmetric if $A = A^T$ and positive-definite if $x^T A x > 0$ for all nonzero $x \in \mathbb{R}^n$. We denote the space of all $n \times n$ symmetric, positive-definite matrices as $PD(n)$. Diffusion tensors are thus elements of $PD(3)$, and structure tensors for 2D images are elements of $PD(2)$. The tangent space of $PD(n)$ at any point can be identified with the space of $n \times n$ symmetric matrices, $Sym(n)$. Given a point $p \in PD(n)$ and a tangent vector X, the Riemannian exponential map is given by

$$\mathrm{Exp}_p(X) = p^{\frac{1}{2}} \exp(\Sigma) p^{\frac{1}{2}T}, \quad \Sigma = p^{-\frac{1}{2}} X p^{-\frac{1}{2}T}, \quad (7)$$

where $\exp(\Sigma)$ is the matrix exponential and can be computed by exponentiating the eigenvalues of Σ, since it is symmetric. Likewise, the Riemannian log map between two
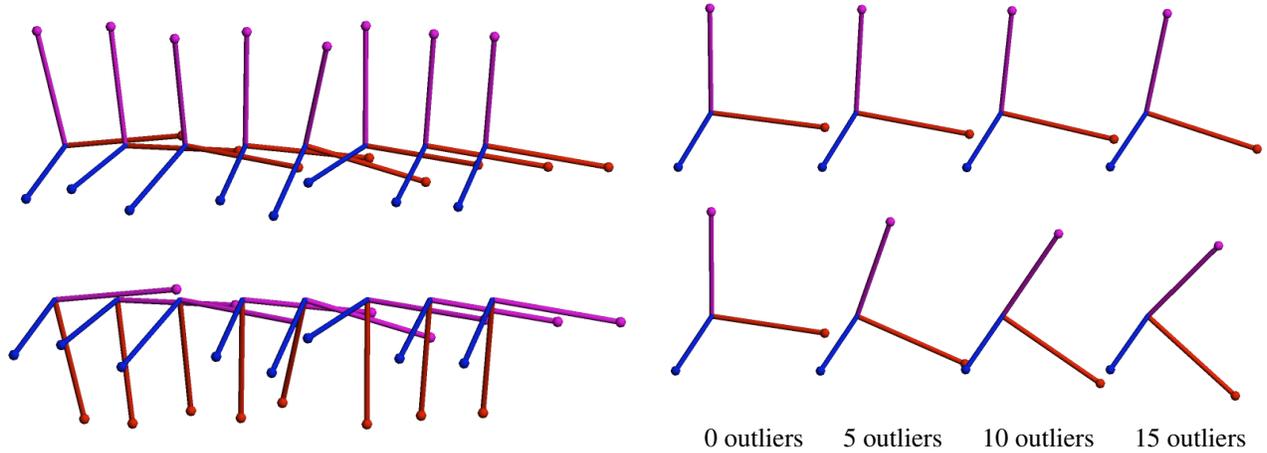
Figure 2. Comparison of the geometric median and Fréchet mean for 3D rotations. Eight rotations from the original dataset (top left). Eight rotations from the outlier set (bottom left). The geometric median results with 0, 5, 10, and 15 outliers (top right). The Fréchet mean results for the same data (bottom right).
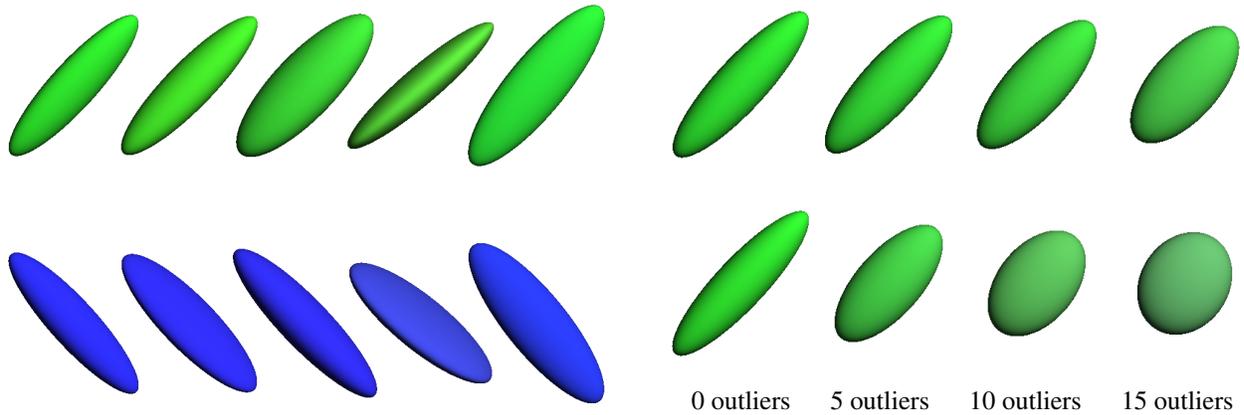

Figure 3. Comparison of the geometric median and Fréchet mean for 3D tensors. Five tensors from the original dataset (top left). Five tensors from the outlier set (bottom left). The geometric median results with 0, 5, 10, and 15 outliers (top right). The Fréchet mean results for the same data (bottom right).

points $p, q \in PD(n)$ is given by

$$\mathrm{Log}_p(q) = p^{\frac{1}{2}} \log(\Lambda) p^{\frac{1}{2}T}, \quad \Lambda = p^{-\frac{1}{2}} q p^{-\frac{1}{2}T}, \quad (8)$$

where $\log(\Lambda)$ is the matrix logarithm, computed by taking the log of the eigenvalues of Λ, which is well defined in the case of positive definite symmetric matrices.

As in the rotations example, we generated 20 random tensors as the image under the exponential map of Gaussian random tangent vectors. The mean was a tensor with eigenvalues $\lambda_1 = 4$ and $\lambda_2 = \lambda_3 = 1$. Next, sets of 5, 10, and 15 outliers were randomly generated in the same fashion with a mean tensor perpendicular to the original group. The standard deviation of both groups was $\sigma = 0.2$. The Fréchet mean and geometric median were computed for the tensor dataset including 0, 5, 10, and 15 outliers, and the results

are shown in Figure 3. Again, convergence of the geometric median took less than 10 iterations in a fraction of a second. The tensors in Figure 3 are colored based on the orientation of the major eigenvector (green = original orientation, blue = outlier orientation) and with color modulated by the fractional anisotropy [2], i.e., more anisotropic tensors are more brightly colored. The geometric median retains the directionality and anisotropy of the original data, unlike the mean, which becomes more isotropic in the presence of outliers. This situation is common in DT-MRI, where adjacent white matter tracts may pass perpendicular to each other. In such cases, the geometric median would be a more appropriate local statistic than the mean to avoid contamination from tensors of a neighboring tract.
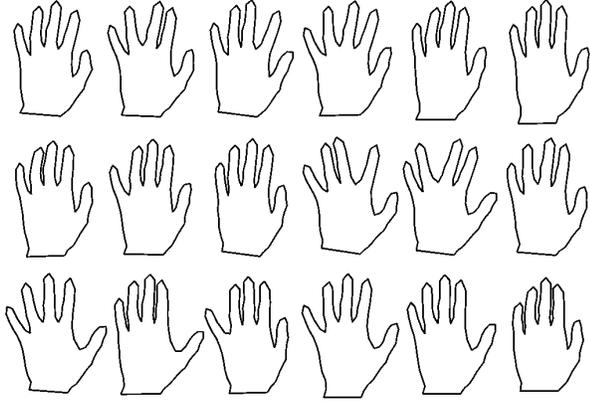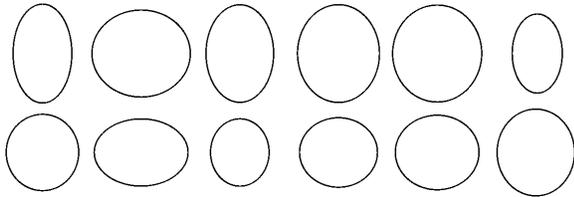
Figure 4. The original dataset of 18 hand shapes.

Figure 5. The 12 outlier shapes.

5.3. Planar Shapes

One area of computer vision that finds the most widespread use of Riemannian geometry is the analysis of shape. Dating back to the groundbreaking work of Kendall [19] and Bookstein [5], modern shape analysis is concerned with the geometry of objects that is invariant to rotation, translation, and scale. This typically results in representing an object's shape as a point in a nonlinear Riemannian manifold, or *shape space*. Recently, there has been a great amount of interest in Riemannian shape analysis, and several shape spaces for 2D and 3D objects have been proposed [14, 16, 20, 26, 31, 35].

An elementary tool in shape analysis is the computation of a mean shape, which is useful as a *template*, or representative of a population. The mean shape is important in image segmentation using deformable models [10], shape clustering, and retrieval from shape databases [32]. The mean shape is, however, susceptible to influence from outliers, which can be a concern for databases of shapes extracted from images. We now present an example showing the robustness of the geometric median on shape manifolds. We chose to use the Kendall shape space as an example, but the geometric median computation is applicable to other shape spaces as well.

We first provide some preliminary details of Kendall's shape space [19]. A configuration of k points in the 2D plane is considered as a complex k-vector, $z \in \mathbb{C}^k$. Remov-

ing translation, by requiring the centroid to be zero, projects this point to the linear complex subspace $V = \{z \in \mathbb{C}^k : \sum z_i = 0\}$, which is equivalent to the space $\mathbb{C}^{k-1}$. Next, points in this subspace are deemed equivalent if they are a rotation and scaling of each other, which can be represented as multiplication by a complex number, $\rho e^{i\theta}$, where ρ is the scaling factor and θ is the rotation angle. The set of such equivalence classes forms the complex projective space, $\mathbb{C}P^{k-2}$. As Kendall points out, there is no unique way to identify a shape with a specific point in complex projective space. However, if we consider that the geometric median only require computation of exponential and log maps, we can compute these mappings relative to the base point, *which requires no explicit identification of a shape with $\mathbb{C}P^{k-2}$*.

Thus, we think of a centered shape $x \in V$ as representing the complex line $L_x = \{z \cdot x : z \in \mathbb{C}\backslash\{0\}\}$, i.e., L_x consists of all point configurations with the same shape as x. A tangent vector at $L_x \in V$ is a complex vector, $v \in V$, such that $\langle x, v \rangle = 0$. The exponential map is given by rotating (within V) the complex line L_x by the initial velocity v, that is,

$$\text{Exp}_x(v) = \cos\theta \cdot x + \frac{\|x\|\sin\theta}{\theta} \cdot v, \quad \theta = \|v\|. \quad (9)$$

Likewise, the log map between two shapes $x, y \in V$ is given by finding the initial velocity of the rotation between the two complex lines L_x and L_y. Let $\pi_x(y) = x \cdot \langle x, y \rangle / \|x\|^2$ denote the projection of the vector y onto x. Then the log map is given by

$$\text{Log}_x(y) = \frac{\theta \cdot (y - \pi_x(y))}{\|y - \pi_x(y)\|}, \quad \theta = \arccos\frac{|\langle x, y \rangle|}{\|x\|\|y\|}. \quad (10)$$

Notice that we never explicitly project a shape onto $\mathbb{C}P^{k-2}$. This has the effect that shapes computed via the exponential map (9) will have the same orientation and scale as the base point x. Also, tangent vectors computed via the log map (10) are valid only at the particular representation x (and not at a rotated or scaled version of x). This works nicely for our purposes and implies that the geometric median shape resulting from (6) will have the same orientation and scale as the intialization shape, m_0.

To test the robustness of the geometric median in Kendall shape space, we used the classic hand outlines from [10]. This data, shown in Figure 4, consists of 18 hand shapes, each with 72 points. We then generated a set of 12 ellipses as outliers. Each ellipse was generated as $(a\cos(\theta_k), b\sin(\theta_k))$, where a, b are two uniformly random numbers in $[0.5, 1]$ and $\theta_k = k\pi/36, k = 0, \ldots, 71$. We computed the Fréchet mean and geometric median for the hand data with 0, 2, 6, and 12 outliers included, corresponding to 0%, 10%, 25%, and 40% outliers, respectively. Both the mean and geometric median computations converge in
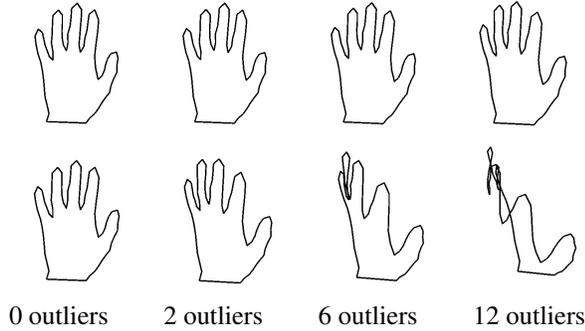
| 0 outliers | 2 outliers | 6 outliers | 12 outliers |

Figure 6. The gometric median shape (top row) from the hand database with 0, 2, 6, and 12 outliers included. The Fréchet mean shape (bottom row) using the same data.

under 15 iterations, running in less than a second for each of the cases. The results are shown in Figure 6. With enough outliers the Fréchet mean is unrecognizable as a hand, while the geometric median is very stable even with 40% outliers. To ensure that both the Fréchet mean and the geometric median computations were not caught in local minima, we initialized both algorithms with several different data points, including several of the outlier shapes. In each case the Fréchet mean and geometric median converged to the same results as shown in Figure 6.

6. Conclusion and Discussion

In this paper we extended the notion of the geometric median, a robust estimator of centrality, to manifold-valued data. We proved that the geometric median exists and is unique for any nonpositively curved manifold and under certain conditions for positively curved manifolds. Generalizing the Weiszfeld algorithm, we introduced a procedure to find the Riemannian geometric median and proved that it converged on positively curved manifolds. Applications to the 3D rotation group, tensor manifolds, and planar shape spaces were presented with comparisons to the Fréchet mean.

Since the area of robust estimation on manifolds is largely unexplored, there are several exciting opportunities for future work. Least squares estimators of the spread of the data have been extended to manifolds via tangent space covariances [28] and principal geodesic analysis (PGA) [14]. Noting that the median is an example of an L^1 M-estimator, the techniques presented in this paper can be applied to extend notions of robust covariances and robust PCA to manifold-valued data. Other possible applications of the Riemannian geometric median include robust clustering on manifolds, filtering and segmentation of manifold-valued images (e.g., images of tensor or directional data).

Appendix

Here we outline the convexity properties of the Riemannian distance function. Our argument follows along the same lines as Karcher [18], who proves the convexity of the squared distance function. Let U be a convex subset of a manifold M. Let $\gamma : [a, b] \to U$ be a geodesic and consider the variation of geodesics from $p \in U$ to γ given by $c(s, t) = \text{Exp}_p(s \cdot \text{Log}_p(\gamma(t)))$. Denote $c' = (d/ds)c(s, t)$ and $\dot{c} = (d/dt)c(s, t)$. (Readers familiar with Jacobi fields will recognize that $\dot{c}$ is a family of Jacobi fields.) The second derivative of the distance function is given by

$$
\frac{d^2}{dt^2} d(p, \gamma(t)) =
$$
$$
\frac{\langle \dot{c}(1, t), (D/ds)\dot{c}(1, t) \rangle}{d(p, \gamma(t))} - \frac{\langle \dot{c}(1, t), c'(1, t) \rangle^2}{d(p, \gamma(t))^3}. \quad (11)
$$

When $\dot{c}(1, t)$ is tangential to $\gamma(t)$, i.e., γ is a geodesic towards (or away from) p, we can easily see that $\frac{d^2}{dt^2} d(p, \gamma(t)) = 0$. Now let $\dot{c}^\perp(1, t)$ be the component of $\dot{c}(1, t)$ that is normal to $\gamma(t)$. We use the following result from [18].

Lemma 1. *If the sectional curvature of M is bounded above by $\Delta > 0$ and $\text{diam}(U) < \pi/(2\sqrt{\Delta})$, then $\langle \dot{c}^\perp(1, t), (D/ds)c'(1, t) \rangle > 0$. If M has nonpositive curvature ($\Delta \leq 0$), then the result holds with no restriction on the diameter of U.*

Along with $\langle \dot{c}^\perp(1, t), c'(1, t) \rangle = 0$, Lemma 1 implies that $\frac{d^2}{dt^2} d(p, \gamma(t))$ is strictly positive when $\text{Log}_{\gamma(t)}(p)$ is not tangential to $\gamma(t)$.

References

[1] C. Bajaj. The algebraic degree of geometric optimization problems. *Discrete and Computational Geometry*, 3:177–191, 1988.

[2] P. J. Basser, J. Mattiello, and D. L. Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysics Journal*, 66:259–267, 1994.

[3] P. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53:221–225, 2005.

[4] J. Bigun, G. Granlund, and J. Wiklund. Multidimensional orientation estimation with application to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, 1991.

[5] F. L. Bookstein. Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1(2):181–242, 1986.

[6] P. Bose, A. Maheshwari, and P. Morin. Fast approximations for sums of distances, clustering and the fermat–weber problem. *Comput. Geom. Theory Appl.*, 24(3):135–146, 2003.

[7] S. R. Buss and J. P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics*, 20(2):95–126, 2001.

[8] R. Chandrasekaran and A. Tamir. Algebraic optimization: The Fermat-Weber problem. *Mathematical Programming*, 46:219–224, 1990.

[9] J. Cheeger and D. G. Ebin. *Comparison Theorems in Riemannian Geometry*. North-Holland, 1975.

[10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Comp. Vision and Image Understanding*, 61(1):38–59, 1995.

[11] U. Eckhardt. Weber's problem and Weiszfeld's algorithm in general spaces. *Mathematical Programming*, 18:186–196, 1980.

[12] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[13] P. T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors. In *Proceedings of ECCV Workshop on Computer Vision Approaches to Medical Image Analysis*, 2004.

[14] P. T. Fletcher, C. Lu, and S. Joshi. Statistics of shape via principal geodesic analysis on Lie groups. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 95–101, 2003.

[15] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10(3):215–310, 1948.

[16] U. Grenander and D. M. Keenan. On the shape of plane images. *SIAM J. Appl. Math.*, 53:1072–1094, 1991.

[17] P. J. Huber. *Robust Statistics*. John Wiley, New York, 1981.

[18] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. on Pure and Appl. Math.*, 30:509–541, 1977.

[19] D. G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:18–121, 1984.

[20] E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[21] H. W. Kuhn. A note on Fermat's problem. *Mathematical Programming*, 4:98–107, 1973.

[22] H. W. Kuhn and R. E. Kuenne. An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. *J. Regional Sci.*, 4:21–34, 1962.

[23] N. Lepore, C. A. Brun, M.-C. Chiang, Y.-Y. Chou, R. A. Dutton, K. M. Hayashi, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. Thompson. Multivariate statistics of the Jacobian matrices in tensor based morphometry and their application to HIV/AIDS. In *MICCAI*, pages 191–198, 2006.

[24] H. P. Lopuhaa and P. J. Rousseeuw. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, 19(1):229–248, 1991.

[25] P. Meer, D. Mintz, D. Y. Kim, and A. Rosenfeld. Robust regression methods in computer vision: a review. *International Journal of Computer Vision*, 6:59–70, 1991.

[26] P. W. Michor and D. Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc.*, 8:1–48, 2006.

[27] L. M. Ostresh. On the convergence of a class of iterative methods for solving the Weber location problem. *Operations Research*, 26:597–609, 1978.

[28] X. Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1), 2006.

[29] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 61(1):41–66, 2006.

[30] Y. Rathi, A. Tannenbaum, and O. Michailovich. Segmenting images on the tensor manifold. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[31] E. Sharon and D. Mumford. 2d-shape analysis using conformal mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 350–357, 2004.

[32] A. Srivastava, S. Joshi, W. Mio, and X. Liu. Statistical shape analysis: clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):590–602, 2005.

[33] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.

[34] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Math. J.*, 43:355–386, 1937.

[35] L. Younes. Computable elastic distances between shapes. *SIAM J. Appl. Math.*, 58:565–586, 1998.